International Journal of Electronics, Communication
& Instrumentation Engineering Research and
Development (IJECIERD)
ISSN(P): 2249-684X; ISSN(E): 2249-7951
Vol. 4, Issue 3, Jun 2014, 33-40
© TJPRC Pvt. Ltd.

TRANS STELLAR
• Journal Publications • Research Consultancy

SALIENCY BASED ADAPTIVE IMAGE SIGNATURE USING BINARY HOLISTIC IMAGE DESCRIPTOR: A SURVEY

RAJESH MATE[1], PRAVIN DERE[2] & SANJAY M. HUNDIWALE[3]

[1]M.E Student, Department of EXTC, ARMIET College of Engineering, Sapgaon, Mumbai, Maharashtra, India

[2]Associate Professor, Department of Electronics, Terna College of Engineering, Nerul, Mumbai, Maharashtra India

[3]Associate Professor, Department of EXTC, ARMIET College of Engineering, Sapgaon, Mumbai Maharashtra, India

ABSTRACT

Finding all objects in a scene and separating them from the background is known as figure-ground separation. For a figure ground separation problem Introduce a simple image descriptor referred to as the image signature. Then show, within the theoretical framework of sparse signal mixing, that this quantity spatially approximates the foreground of an image. Thus experimentally investigate whether this approximate foreground overlaps with visually conspicuous image locations by developing a saliency algorithm based on the image signature. This saliency algorithm predicts human fixation points best among benchmark data set and does so in much shorter running time. In a related experiment, Thus demonstrate with a change blindness data set that the distance between images induced by the image signature is closer to human perceptual distance than can be achieved using other saliency algorithms, pixel-wise, or GIST descriptor methods.

KEYWORDS: Saliency Based Adaptive Image Signature, Saliency Algorithms, Pixel-Wise, GIST

INTRODUCTION

Salient areas in natural scenes are generally regarded as areas which the human eye will typically focus on, and finding these areas is the key step in object detection. In computer vision, many models have been proposed to simulate the behavior of eyes such as Saliency Toolbox (STB), Neuromorphic Vision Toolkit (NVT), and others, but they demand high computational cost and computing useful results mostly relies on their choice of parameters. Although some region-based approaches were proposed to reduce the computational complexity of feature maps, these approaches still were not able to work in real time. Thus provide an approach to the figure-ground separation problem using a binary, holistic image descriptor called the “image signature.”

It is defined as the sign function of the Discrete Cosine Transform (DCT) of an image. Then demonstrate, this simple descriptor preferentially contains information about the foreground of an image—a property which believes underlies the usefulness of this descriptor for detecting salient image regions. Then formulate the figure-ground separation problem in the framework of sparse signal analysis. Prove that the Inverse Discrete Cosine Transform (IDCT) of the image signature concentrates the image energy at the locations of a spatially sparse foreground, relative to a spectrally sparse background. Then, demonstrate this phenomenon on synthetic images with sparse foregrounds much weaker in intensity than the complex background pattern.

Two experiments are presented to quantify the relationship between the image signature and human visual attention. Demonstrate that a saliency map derived from the image signature outperforms many leading saliency algorithms on a benchmark data set of eye-movement fixation points. Then introduce reaction time data collected from nine subjects in a change blindness experiment. And show that the distance between images induced by the image signature most closely matches the perceptual distance between images inferred from these data among competing measures derived from other saliency algorithms, the GIST descriptor, and simpler pixel measures.

PROBLEM DEFINITION

Most traditional object detectors need training in order to detect specific object categories but human vision can focus on general salient objects rapidly in a clustered visual scene without training because of the existence of visual attention. Therefore, humans can easily deal with general object detection well, which is becoming an intriguing subject for more and more research. Salient object detection is an important technique for many content-based applications, but it becomes a challenging work when handling the cluttered saliency maps, which cannot completely highlight salient object regions and cannot suppress background regions. The problem of finding all objects in a scene and separating them from the background is known as figure-ground separation. The human brain can perform this separation very quickly and doing so on a machine remains a major challenge for engineers and scientists. The problem is closely related to many of the core applications of machine vision, including scene understanding, content-based image retrieval, object recognition, and tracking.

LITERATURE SURVEY

In 1890, James suggested that visual attention operates like a “spotlight” that can move around the visual field. What attracts people’s attention? Tresiman proposed the famous feature integration theory (FIT) which described visual attention as having two stages. A set of basic visual features, such as color, motion and edges, is processed in parallel at the preattentive stage. And then, in the limited-capacity process stage, the visual cortex performs other more complex operations like face recognition and others. A master map or a saliency map is computed to indicate the locations of salient areas. Distinctive features (e.g., luminous color, high velocity motion, and others) will “pop out” automatically in the preattentive stage, and then the salient areas become the object candidates.

- **An Application to Saliency-Based Invariant Image Feature Construction**

In May 2008 Huicheng Zheng, Member, IEEE, Grégoire Lefebvre, and Christophe Laurent implemented adaptive-subspace self-organizing map (ASSOM) and it is useful for saliency based invariant feature generation and visualization. However, the learning procedure of the ASSOM is slow. But In this paper, two fast implementations of the ASSOM are proposed to boost ASSOM learning based on insightful discussions of the basis rotation operator of ASSOM. They investigate the objective function approximately maximized by the classical rotation operator. They explore a sequence of two schemes to apply the proposed ASSOM implementations to saliency-based invariant feature construction for image classification.

- **Camera Motion-Based Analysis of User Generated Video**

In Jan 2010 Golnaz Abdollahian propose a system for the analysis of user generated video (UGV). UGV often has a rich camera motion structure that is generated at the time the video is recorded by the person taking the video, i.e., the “camera person.” They exploit this structure by defining a new concept known as camera view for temporal segmentation of UGV. The segmentation provides a video summary with unique properties that is useful in applications such as video annotation. Camera motion is also a powerful feature for identification of key frames and regions of interest (ROIs) since it is an indicator of the camera person's interests in the scene and can also attract the viewers' attention. They propose a new location-based saliency map which is generated based on camera motion parameters. This map is combined with other saliency maps generated using features such as color contrast, object motion and face detection to determine the ROIs.

- **A Novel Multi Resolution Spatiotemporal Saliency Detection Model and its Applications in Image and Video Compression**

In Jan 2010 Chenlei Guo, Student Member, IEEE, and Liming Zhang, Senior Member, IEEE presents a quaternion representation of an image which is composed of intensity, color, and motion features. Based on the principle of phase spectrum of Fourier transform (PFT) model, a novel multi resolution spatiotemporal saliency detection model called phase spectrum of quaternion Fourier transform (PQFT) is proposed in this paper to calculate the spatiotemporal saliency map of an image by its quaternion representation.

- **Saliency and Gist Features for Target Detection in Satellite Images**

In July 2011 Zhicheng Li and Laurent Itti explores an automatic approach to detect and classify targets in high-resolution broad-area satellite images, which relies on detecting statistical signatures of targets, in terms of a set of biologically-inspired low-level visual features. Broad-area images are cut into small image chips, analyzed in two complementary ways: "attention/saliency" analysis exploits local features and their interactions across space, while "gist" analysis focuses on global non spatial features and their statistics. Both feature sets are used to classify each chip as containing target(s) or not, using a support vector machine.

METHODOLOGY

Our aim is to design an approach to the figure-background separation problem using a binary, holistic image descriptor called the "image signature". Thus Obtain silency map using various methods We compare the silency maps formed from image signature to following silency algorithms i.e. GBVS, SUN, Itti etc.

This descriptor preferentially contains information about the foreground of an image—a property of this descriptor for detecting salient image regions **Image signature.** We begin by considering gray-scale images which exhibit the following structure:

$$\mathbf{x = f + b, x, f, b \in IR^N} \quad (1)$$

where f represents the foreground and is assumed to be sparsely supported in spatial(time) domain. b represents the background and is assumed to be sparsely supported in the basis of the Discrete Cosine Transform. In other words, both f and b have only a small number of nonzero components.

For the problem of figure-ground separation, we are only interested in the spatial support of f (the set of pixels for which f is nonzero) we can approximately **isolate** the support of f by taking the **sign** of the mixture signal x in the transformed domain (dct) then inversely transform it back into the spatial domain, i.e., by computing the reconstructed image Formally, the image signature is defined as

Image Signature (x) = sign (DCT (X))

FLOWCHART

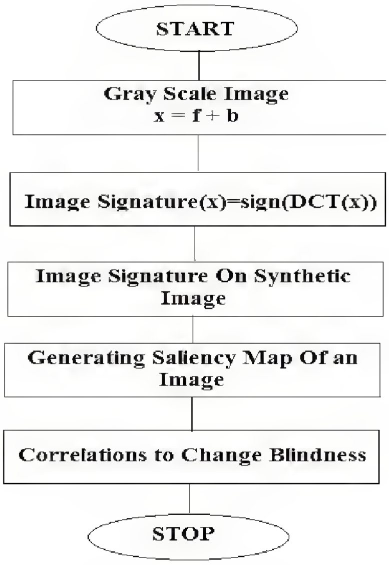
START
Gray Scale Image
x = f + b
Image Signature(x)=sign(DCT(x))
Image Signature On Synthetic Image
Generating Saliency Map Of an Image
Correlations to Change Blindness
STOP

Figure 1

IMAGE DESCRIPTORS

Visual descriptors or **image descriptors** are descriptions of the visual features of the contents in **images, videos, algorithms, or applications.** They describe elementary characteristics such as the **shape, color, texture or the motion**. These descriptors have a good knowledge of the objects and events found in a video, image. Visual descriptors are divided in two main groups: **1) General information descriptors**: they contain low level descriptors which give a description about colour, shape, regions, textures & motion. **2) Specific domain information descriptors**: they give information about objects and events in the scene. example - face reorganization.

ADVANTAGES OF IMAGE SIGNATURE

1) The image signature discards amplitude information across the entire frequency spectrum, **storing only the sign of each DCT component. 2**) The image signature is thus **very compact**, with a single bit per component.

SILENCY MAP

If an image foreground is **clearly visible** relative to its background, then we can form a **saliency map.** Saliency map shows the shape of foreground support of the image. **It is the saliency which drives our attention.** Various algorithms can be used to calculate saliency map. for example RGB signature, lab signature, GBVS, Sun, Itti Each saliency map generated by the algorithm is threshold and then considered as a binary classifier. The point in the image f(x, y)> T is called object point(foreground) otherwise it is called background point. i.e. Any pixel having intensity value more than threshold value is considered as foreground, otherwise part of background.

GRAPH BASED VISUAL SALIENCY (GBVS)

A method of computing bottom-up saliency maps which shows a remarkable consistency with the attention deployment of human subjects. It highlights handful of ‘significant’ locations where the image is ‘informative’ according to some criterion, e.g. human Fixation. In this method first computing feature maps (step1), e.g. by linear filtering followed by some elementary nonlinearity "Activation" (step2), form an "activation map" (or maps) using the feature vectors "**normalization and combination" (s3) step.** Aim of any saliency algorithm: concentrating activation into a few key locations.

EXPECTED RESULTS

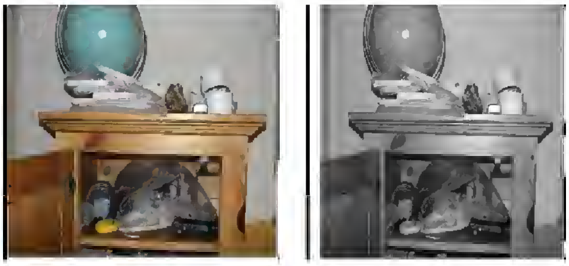

Figure 2: Original Image   Saliency Map (GBVS)

ADVANTAGES OF GBVS

1) GBVS predicts human fixations more reliably than the standard algorithms. 2) GBVS robustly highlights salient regions, even far away from object borders. 3) GBVS promotes higher saliency values in the center of the image plane.

SUN Saliency Using Natural Statistics

Natural images are distinctive, because they contain particular types of structure. Natural images are not Gaussian; natural scenes are statistically quite different from white noiseless storage space required. If every image consisted of a uniform grey, then just one number would specify the whole image: its grey level. Natural scenes possess scale invariance

It is a Bayesian framework (based on Baye's Theorem) from which bottom-up saliency emerges naturally as the self-information of visual features; overall saliency emerges as the point wise mutual information between the features and the target when searching for a target.

Bottom up Saliency: Driven by scene features, fast!

Top Down Saliency: Driven by will control, task oriented, Slow.

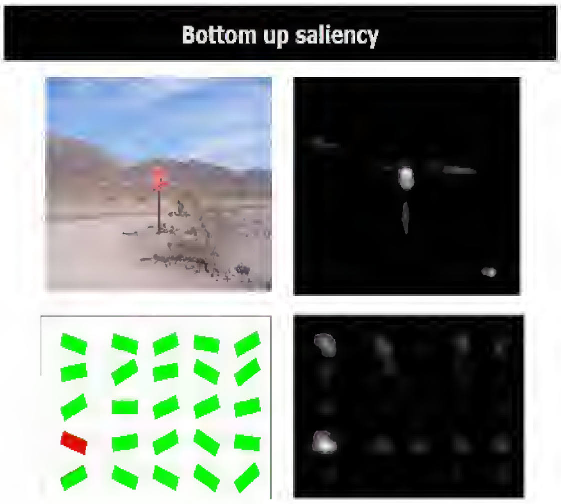

Figure 3: Bottom up Saliency

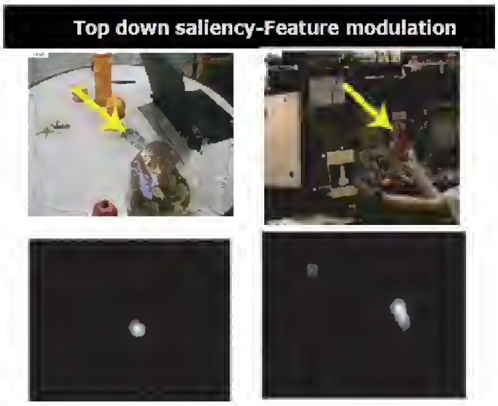

Figure 4: Top down Saliency Feature Modulation

EXPECTED RESULTS

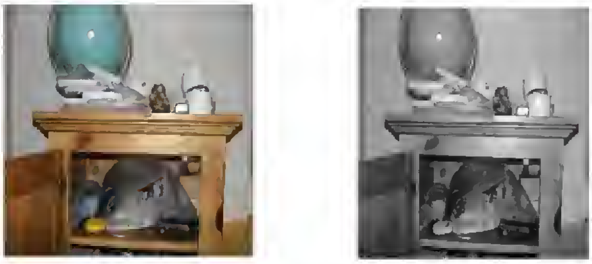

Figure 5: Original Image Sun Saliency Map

In sun method saliency is computed locally, a normalization scheme based on local maxima ("max-ave"), which is consistent with the neuroanatomy of the early visual system and results in an efficient algorithm with few free parameters. SUN is likely to function best for short durations when the task is simple.

PROPOSED METHOD

We will use ***Quad tree decomposition*** technique here so that image is divided in four sections. The partition continues till the pixel values are different. For image section if the pixel values are same no further decomposition is done.

Advantages

- Process becomes ***Adaptive.***
- No of descriptors required reduces significantly.
- No. of computations reduces. Algorithm becomes faster.

EXPECTED RESULTS

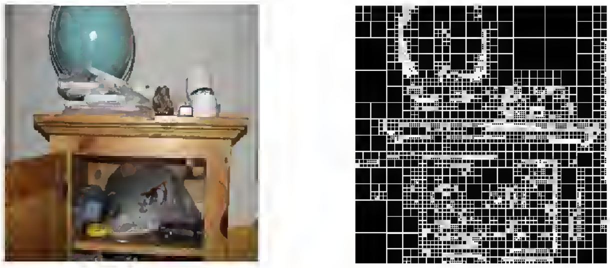

Figure 6: Original Image Quad Tree Decomposition

A **quad tree** is a tree data structure in which each internal node has exactly four children. Quad trees are most often used to partition a two dimensional space by recursively subdividing it into four quadrants or regions. The regions may be square or rectangular, or may have arbitrary shapes. This data structure was named a quad tree by Raphael Finkel and J.L. Bentley in 1974. A similar partitioning is also known as a *Q-tree*. All forms of Quad trees share some common features:

They decompose space into adaptable cells Each cell (or bucket) has a maximum capacity. When maximum capacity is reached, the bucket splits The tree directory follows the spatial decomposition of the Quad tree.

Types

Quad trees may be classified according to the type of data they represent, including areas, points, lines and curves. Quad trees may also be classified by whether the shape of the tree is independent of the order data is processed. Some common types of quad trees are:

The Region Quadtree

The region quadtree represents a partition of space in two dimensions by decomposing the region into four equal quadrants, sub quadrants, and so on with each leaf node containing data corresponding to a specific sub region. Each node in the tree either has exactly four children, or has no children (a leaf node). The region quadtree is a type of trie. A region quadtree with a depth of n may be used to represent an image consisting of $2^n \times 2^n$ pixels, where each pixel value is 0 or 1. The root node represents the entire image region. If the pixels in any region are not entirely 0s or 1s, it is subdivided. In this application, each leaf node represents a block of pixels that are all 0s or all 1s.

A region quadtree may also be used as a variable resolution representation of a data field. For example, the temperatures in an area may be stored as a quadtree, with each leaf node storing the average temperature over the sub region it represents. If a region quadtree is used to represent a set of point data (such as the latitude and longitude of a set of cities), regions are subdivided until each leaf contains at most a single point.

Point Quadtree

The point quadtree is an adaptation of a binary tree used to represent two dimensional point data. It shares the features of all quadtrees but is a true tree as the center of a subdivision is always on a point. The tree shape depends on the order data is processed. It is often very efficient in comparing two dimensional ordered data points, usually operating in O (log n) time.

Node Structure for a Point Quadtree

A node of a point quadtree is similar to a node of a binary tree, with the major difference being that it has four pointers (one for each quadrant) instead of two ("left" and "right") as in an ordinary binary tree. Also a key is usually decomposed into two parts, referring to x and y coordinates. Therefore a node contains following information: 4 Pointers: quad['NW'], quad['NE'], quad['SW'], and quad['SE'] point; which in turn contains: 1)key; usually expressed as x, y coordinates 2)value; for example a name

Edge Quadtree

Edge quadtrees are specifically used to store lines rather than points. Curves are approximated by subdividing cells to a very fine resolution. This can result in extremely unbalanced trees which may defeat the purpose of indexing.

Polygonal Map Quadtree

The Polygonal Map Quad tree (or PM Quad tree) is a variation of quad trees which are used to store collections of polygons that may be degenerate (meaning that they have isolated vertices or edges).

CONCLUSIONS

We introduced the image signature as a simple yet powerful descriptor of natural scenes. We proved on the basis of theoretical arguments that this descriptor can be used to approximate the spatial location of a sparse foreground hidden in a spectrally sparse background. predicting them better than leading saliency algorithms at a fraction of the computational cost. We also provided hange blindness experiment in which the perceptual distance between slightly different images was predicted most accurately by the image signature descriptor. In future we are investigate experimentally the same.

Future Scope

Thus in future we are investigate more number of saliency map of image signature which have better performance than quad tree and other saliency map descriptor.

REFERENCES

1. N. Bruce and J. Tsotsos, “Saliency Based on Information Maximization,” Proc. Advances in Neural Information Processing Systems, pp. 155-162, 2006.
2. Oliva and A. Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope,” Int’l J. Computer Vision, vol. 42, no. 3, pp. 145-175, 2001.
3. H. Zhou, H. Friedman, and R. von der Heydt, “Coding of Border Ownership in Monkey Visual Cortex,” J. Neuroscience, vol. 20, no. 17, pp. 6594-6611, 2000.
4. E. Cande`s, X. Li, Y. Ma, and J. Wright, “Robust Principal Component Analysis?” Arxiv preprint arXiv: 0912.3599, 2009.
5. X. Hou and L. Zhang, “Saliency Detection: A Spectral Residual Approach,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.
6. Oppenheim and J. Lim, “The Importance of Phase in Signals,” Proc. IEEE, vol. 69, no. 5, pp. 529-541, May 1981.
7. M. Hayes, J. Lim, and A. Oppenheim, “Signal Reconstruction from Phase or Magnitude,” IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 28,no. 6, pp. 672-680, 1980.
8. L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.